

## АПОСТЕРИОРЕН АНАЛИЗ НА ДИДАКТИЧЕСКИ ТЕСТ ЧРЕЗ СПЕЦИАЛИЗИРАН СОФТУЕР<sup>1</sup>

**гл. ас. Евгения Ангелова\*, докторант Радослав Радев\*\***

\* катедра Компютърни технологии,  
ФМИ при ПУ „П. Хилендарски“  
гр. Пловдив, бул. България, 236  
evgang@uni-plovdiv.org

\*\* катедра Компютърни технологии,  
ФМИ при ПУ „П. Хилендарски“  
гр. Пловдив, бул. България, 236  
radoslav@kodar.net

Дидактическият тест е измерителен инструмент за проверка на степента на усвоеност на знанията и уменията на обучаемите. Включените в теста задачи и тестът като цяло трябва да отговарят на определени критерии. Предлаганият програмен продукт улеснява апостериорния анализ на дидактическия тест, като на базата на въведени отговори се изследват различни характеристики както на отделните тестови задачи (анализ на: трудност, дискриминативна сила, дистрактори), така и на теста като цяло (надеждност, валидност). Наличие на неприемливи стойности на показателите предполага промяна на съответните въпроси.

**Ключови думи:** анализ на дидактически тест, индекс на трудност, дискриминативна сила, анализ на дистрактори.

**1. Въведение.** Една от задачите на обучението е проверка и оценка на знанията, уменията и навиците, придобити от учещите се в двустранния процес на взаимодействие между преподавателя и обучаемите. Много често като средство за измерване на постиженията на обучаемите при усвояване на определено учебно съдържание се използват дидактически тестове. Чрез тестовете се установява: индивидуалният статус на тестируания индивид или група в сравнение с постиженията на други учещи се (нормативни тестове [1, с. 295]); нивото на постигане на целите и задачите на методическата система по отношение на съответния етап на усвоеност на знанията и уменията (критериални тестове). Тестът е последователност от въпроси в определена област. Класическата класификация на тестовите въпроси, към която се придържат Торндайк, Рютер и Хербиг, е според очаквания отговор: задачи със структуриран отговор (закрит тип) – избира се верен отговор от ограничено множество алтернативи; задачи със свободен отговор (отворен тип) – обучаемите сами изграждат своя отговор [2, с. 310]. Чрез задачите от закрит тип обучаемият оперира с една напълно определена структурирана ситуация; оценката е максимално обективизирана; може да се покрие напълно учебното съдържание и да се постигне съответствие с целите на измерването. Проверката е сравнително бърза, поради което се предпочита от преподавателите и най-често те сами разработват такива тестове. Тестовете, за да бъдат ефективна форма за оценяване, трябва да са стандартизирани, съпроводени с доказателство, че действително мерят постиженията, за които са предвидени. Това предполага те да притежават определени качества: обективност относно подготовката, провеждането на теста и анализ на резултатите; надеждност; валидност и др. [2, с. 285-290].

**2. Апостериорен анализ на тест.** Разработените тестови задачи се апробират, за да се получат данни и да се изследват качествата на теста след неговото изпробване, преди да се подготви за измерване на постиженията на обучаемите. Анализът се извършва по отношение на характеристиките на тестовите задачи: трудност, дискриминативна сила, анализ на дистракторите.

**Анализ на трудността.** Трудността на тестовата задача се определя чрез т. нар. **индекс на трудност P**: процентът на решените правилно задачата спрямо общия брой тествани лица [2, с. 342-343].

$$P = 100 \frac{N_R}{N}; \text{ където}$$

$N_R$  е броят на тестваните лица, решили вярно задачата;

$N$  е общият брой на всички тествани, решавали задачата.

<sup>1</sup> Тази работа е подпомогната по проект ИС–М–4 на поделение „Научна и приложна дейност“ при Пловдивския университет „Паисий Хилендарски“.

Трудността се тълкува в зависимост от броя на алтернативите за избор [5, с. 66]. При индекс на трудност не повече от 25% на една задача с четири алтернативи тя се счита за много трудна, а ако не е по-малък от 85% – за много лесна, т. е. оптимална задача е с индекс на трудност в граници от 40% до 70%. Това са стойности, които в разработения програмен продукт са параметризирани за ползването му при тестови задачи с различен брой алтернативи.

**Анализ на дискриминативната сила.** Тази стойност показва възможностите на дадена задача да разграничи силните от слабите по постижения лица [2, с. 339]. За целта се определят две групи – „силна“ и „слаба“, като тестираните лица се подредят по общия бал от теста. Според някои автори се вземат първите 27%, получили най-висок резултат, и последните 27% с най-нисък резултат, докато други автори препоръчват в екстремалните групи да се включат първите и последните 33% от работилите теста. Броят на тестираните лица, определящ двете екстремални групи, се пресмята на базата на параметър  $X_t$  в програмния продукт. Висока стойност на характеристиката показва добро разграничаване на „силните“ от „слабите“ тестиращи обучаеми. Измерва се чрез т. нар. **индекс на дискриминативната сила**, който по същество представлява коефициента на корелация между групата на силните и групата на слабите тестиращи лица и се пресмята по формулата [2, с. 339]:

$$DP = \frac{R_U - R_L}{T/2}; \text{ където}$$

$R_U$  е броят на тестираните лица от „силната“ група, които са решили вярно задачата;

$R_L$  е броят на тестираните лица от „слабата“ група, които са решили вярно задачата;

$T$  е общият брой на тестираните лица от двете екстремални групи.

Както е отбелязано в [3, с. 343], една стойност на индекса на дискриминативна сила е приемлива, ако тя е в интервала [+40; +60], а според Р. Ебел и др. една задача следва да се остави без внимание, ако съответният ѝ индекс не е в интервала [+20; +80].

**Анализ на дистракторите.** Терминът дистрактор се използва за назоваване на неправилните алтернативи [5, с. 71]. Анализът на дистракторите на определен въпрос или задача се извършва на базата на три критерия, посочени в [3, с. 344-345]:

а) за повечето задачи посочилите даден дистрактор от слабата група са повече от посочилите го от силната група;

б) всеки дистрактор е посочен и от тестиран от силната група;

в) посочилите дистрактор от силната група не са повече от тези, посочили верен отговор, от същата група.

Наличие на дистрактор, който не е посочен от тестиран както от слабата, така и от силната група, показва, че не е привлякъл вниманието поради няколко причини: почти всички учещи владеят добре материала, тъй като за него е отделено повече време; съществуват двойки алтернативи, за които се знае, че едната е вярна; използвани са невъзможни алтернативи [5, с. 72]. След този анализ е необходимо редактиране на алтернативите на съответните задачи (в примера – Табл. 1, задача 5-г).

Група	В 1				В 2				В 3				В 4				В 5			
	а	б	в	г	а	б	в	г	а	б	в	г	а	б	в	г	а	б	в	г
силна	2	13	1	2	0	2	16	0	1	15	2	0	5	2	3	8	14	3	1	0
слаба	12	1	3	2	5	4	8	1	3	2	9	4	2	6	7	3	6	4	8	0

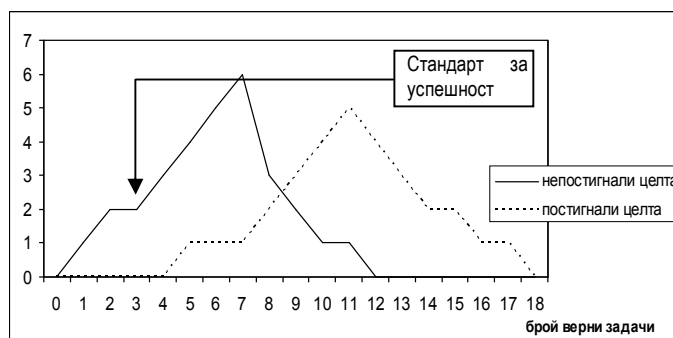
Табл. 1

**Валидност на теста.** В методиката за конструиране на дидактически тестове се изисква изследване на качествата надеждност и валидност. Надеждността на един тест показва неговата точност на измерване [2, с. 342]. Валидността показва степента на точност, с която измерва това, което трябва да измерва. Според [8, с. 98] валидността на един тест е: критериална, съдържателна, конструктивна. Критериалната валидност показва степента на съответствие на дидактическия тест с външен, предварително определен критерий, като в [2, с. 351] са дадени статистически методи за анализ и изследване на тази валидност. При метода на контрастните групи се обособяват: „силна“ група (обучаемите, резултатите на които показват, че са постигнали в известна степен целта на обучението) и „слаба“ група (обучаемите, които не са постигнали тази цел). Въз основа на тези данни се попълва таблицата с честотното разпределение на баловите на тестираните лица от двете групи (Табл. 2) (в примера обучаемите решават 18 тестови задачи, като всяка вярно решена задача се оценява с 1 точка):

Балове	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Слаба група	0	1	2	2	3	4	5	6	3	2	1	1	0	0	0	0	0	0	0
Силна група	0	0	0	0	0	1	1	1	2	3	4	5	4	3	2	2	1	1	0

Табл. 2

На Фиг. 1 са построени полигоните на разпределението за всяка от двете групи при една и съща координатна система.



Фиг. 1

Абсцисата на пресечната точка на двете графики се приема за **стандарт на успешност**. Както се вижда, това е число от интервала (8; 9). При стандарт за успешност осем верни задачи, тестираните от „силната“ група, получили под 8 точки, са общо 3, и ще бъдат категоризирани като „непостигнали целите“. Тестираните от „слабата“ група, получили поне 8 точки (общо 7), се оценяват като „постигнали целта“. При така определеният стандарт общо 10 (3 и 7) тестирани няма да бъдат оценени правилно (16,67%). За приемлива стойност на критериалната валидност се приема резултат, по-голям от 60%.

**3. Възможности на програмния продукт.** Програмният продукт подпомага разработчика на тестови задачи в извършването на апостериорния анализ, за да може тестът да се прилага за измерване на постиженията на обучаеми. На базата на въведените отговори от тестовите задачи се изследват измерителните качества на теста.

Необходимо е въвеждане на изработени тестове, като за всеки тест се въвеждат: брой задачи в теста; брой алтернативи (възможни отговори) на задачите; верните отговори на задачите; отговорите, дадени от тестираните лица.

Програмният продукт позволява изработване на различни справки, като:

- Пресмятане на броя на верните отговори на всеки тестиран и подреждане на резултатите по общ бал;
- Селектиране на първите (с най-висок бал)  $X_f$  и последните (с най-нисък бал)  $X_f$  тестиращи лица за определяне на двете екстремални групи („силна“ и „слаба“);
- Пресмятане **„индекс на трудност“** – броят на решените вярно задачите, разделен на общия брой тестиращи лица; по-висок процент означава по-лесна задача;
- Пресмятане на **„дискриминативна сила“** – разликата от процента на броя на тестираните лица от „силната“ група, които са решили вярно задачите и процента на вярно решените теста от слабата група;
- **Оценка на дистрактори** – програмният продукт предоставя информация за тестовите задачи с неприемливи дистрактори и съставителят на теста да коригира съответните алтернативи.
- **Валидност на теста** – специализираният софтуер визуализира таблицата за честотното разпределение на баловете на тестираните лица от двете екстремални групи, изчертава полигоните на разпределението за броя на верните отговори и посочва стандарта за успешност на теста.

**4. Заключение.** Стандартизирането на теста го прави ефективна форма за оценяване с предимство пред другите методи, които не подлежат на стандартизация. Тестът е широко разпространен и добре разработен инструмент за измерване на постиженията на обучаемите както в училище, така и във висшето образование [6]. Разработеният програмен продукт може да се използва от: учители, които сами разработват тестове; групи за разработка на тестове за входно и изходно ниво в училище; специалисти, занимаващи се с педагогически изследвания. Възможностите на информационните и комуникационни технологии позволяват създаване и на подходящ програмен инструментариум за електронно тестово изпитване и оценяване. Нуждата и някои плюсове от автоматизирани системи за тестово изпитване са

дискутирани в [10], в [9] е направен опит да се даде архитектурно решение за разширяеми изчислителни и в частност някои тестови системи. В [7] е описан разпределен клъстер за електронно тестване – DeTC, снабден с възможности за създаване и оценяване на статични и динамични тестови задачи, за генериране на тестове, за провеждане на групово изпитване и оценяване, което е обективно и експресно [4].

#### **Използвана литература**

- [1] Бижков, Г., Теория и методика на дидактическите тестове, Просвета, София, 1992.
- [2] Бижков, Г., Методология и методи на педагогическите изследвания, Аскони-Издат, София, 1995.
- [3] Бижков, Г., В. Краевски, Методология и методи на педагогическите изследвания, УИ „Св. Климент Охридски“, София, 2002.
- [4] Рахнев А., О. Рахнева, Н. Вълчанов, Приложения на DeTC за изпитване и оценяване в квалификация „Учител по информационни технологии“, Математика и математическо образование, София, 2007, ISBN 978-954-8880-25-1, стр. 397-403.
- [5] Стоименова, Е., Измерителни качества на тестовете, НБУ, София, 2000.
- [6] Brusilovsky P., Philip L. Miller, Web-based Testing for Distance Education, WebNet 1999, pp. 149-155.
- [7] Rahneva, O., A. Rahnev, N. Pavlov, Functional Workflow and Electronic Services In a Distributed Electronic Testing Cluster – DeTC , Proceedings 2nd International Workshop on eServices and eLearning, Otto-von-Guericke Universitaet Magdeburg, 2004, ISBN 3-929757-76-1, pp 147-157.
- [8] Standarts for educational and psychological testing, Prep. by the Committee to Develop Standarts for Educational and Psychological Measurement in Education, Washington, American Psychological Assotiation, 1985.
- [9] Valchanov, N., T. Terzieva, V. Shkurtov, A. Iliev, Architecture of extensible computations driven systems, Сборник доклади на Тридесет и деветата пролетна конференция на Съюза на математиците в България, Албена, 6-10.04.2010, 207-211.
- [10] Илиев, А., Н. Вълчанов, Т. Терзиева. Споделен опит от използване на софтуерна тестова система при провеждане на изпити в курса по Информационно моделиране, сборник с доклади от Юбилейна конференция „Науката, образованието и времето като грижа“, 30.11-01.12.2007 г., Смолян, 77-81.